Reconstructing a multinormal covariance matrix from its spherically truncated projection

Filippo Palombi, Simona Toti and Romina Filippini

Abstract Suppose $X \sim \mathcal{N}_v(0,\Sigma)$ with $v \geq 1$. If the probability density of X is cut off outside a centered Euclidean ball with given radius, the matrix $\mathfrak{S}$ of the second truncated moment differs from Σ. In this talk we discuss the effects of the spherical truncation and the reconstruction of Σ from $\mathfrak{S}$. The latter is achieved numerically thanks to a fixed point iteration. We study the convergence rate of the iteration and propose an acceleration scheme.

Key words: Distributional truncation, covariance reconstruction, fixed point iteration

1 A motivating example

Consider following experimental situation. An accelerator physicist prepares a particle beam with Gaussian transversal profile. The experimenter knows *a priori* the spatial distribution of the beam, i.e. covariance matrix Σ of the two–dimensional coordinates of the particles on a transversal slice of the beam. The beam travels straightforward until it enters a linear coaxial pipeline with circular profile. Part of the beam is absorbed, part propagates within the pipeline. At the end of beam flight the physicist wants to know if the transversal distribution of the particles is changed, due to inter–beam interactions. Accordingly, he measures again the spatial covariance matrix of the two dimensional coordinates. Unfortunately, the pipeline cut off introduced a bias on the covariance matrix, which is therefore not directly comparable with the original one. It is possible to remove the bias? In the sequel we generalize the problem to $v \geq 1$ dimensions and discuss its solution. For a fully detailed discussion, we refer the reader to ref. [2].

Filippo Palombi, Simona Toti and Romina Filippini
Istituto Nazionale di Statistica, Via C. Balbo 16, 00184 Rome – Italy
e-mail: filippo.palombi@istat.it, simona.toti@istat.it, romina.filippini@istat.it

2 Set up of the problem

Let $X \in \mathbb{R}^v$ be a random vector with jointly normal distribution $\mathscr{N}_v(0,\Sigma)$ in $v \geq 1$ dimensions. The probability that X falls within $\mathscr{B}_v(\rho) \equiv \{x \in \mathbb{R}^v : x^{\mathrm{T}}x < \rho\}$ is measured by the Gaussian integral

$$\alpha(\rho;\Sigma) \equiv \mathbb{P}\left[X \in \mathscr{B}_v(\rho)\right] = \frac{1}{(2\pi)^{v/2}|\Sigma|^{1/2}} \int_{\mathscr{B}_v(\rho)} \mathrm{d}^v x \, \mathrm{e}^{-\frac{1}{2}x^{\mathrm{T}}\Sigma^{-1}x}. \tag{1}$$

Since Σ is symmetric positive definite, it has real orthonormal eigenvectors $\Sigma v^{(i)} = \lambda_i v^{(i)}$. Let us denote by $R \equiv \{v_i^{(j)}\}_{i,j=1}^v$ the special orthogonal matrix having these vectors as columns and by $\Lambda \equiv \mathrm{diag}(\lambda) = R^{\mathrm{T}}\Sigma R$ the diagonal counterpart of Σ, being $\lambda = \{\lambda_1, \ldots, \lambda_v\}$. From the invariance of $\mathscr{B}_v(\rho)$ under rotations, it follows that α depends upon Σ just by way of λ. Accordingly, we rename the Gaussian probability content of $\mathscr{B}_v(\rho)$ as

$$\alpha(\rho;\lambda) \equiv \int_{\mathscr{B}_v(\rho)} \mathrm{d}^v x \prod_{m=1}^{v} \delta(x_m,\lambda_m), \quad \delta(y,\eta) = \frac{1}{\sqrt{2\pi\eta}} \exp\left\{-\frac{y^2}{2\eta}\right\}. \tag{2}$$

By contrast, the matrix of the second truncated moments $\mathfrak{S}$ depends actually on both λ and R. An explicit calculation yields

$$\alpha(\rho;\lambda) \cdot \mathfrak{S}_{ij} = \sum_{k,\ell=1}^{v} R_{ki}R_{\ell j} \int_{\mathscr{B}_v(\rho)} \mathrm{d}^v x \, x_k x_l \prod_{m=1}^{v} \delta(x_m,\lambda_m). \tag{3}$$

From eq. (3) it follows that Σ and $\mathfrak{S}$ share R as a common diagonalizing matrix. In other words, if $M \equiv \mathrm{diag}(\mu)$ is the diagonal matrix of the eigenvalues $\mu = \{\mu_1, \ldots, \mu_v\}$ of $\mathfrak{S}$, then $M = R^{\mathrm{T}}\mathfrak{S}R$. Moreover, μ_k is related to λ_k by

$$\mu_k = \lambda_k \frac{\alpha_k}{\alpha}, \quad \alpha_k(\rho;\lambda) \equiv \int_{\mathscr{B}_v(\rho)} \mathrm{d}^v x \, \frac{x_k^2}{\lambda_k} \prod_{m=1}^{v} \delta(x_k,\lambda_k); \quad k = 1,\ldots,v. \tag{4}$$

Eqs. (4) constitute a system of non–linear integral equations, with the eigenvalues λ_k as unknown variables and the eigenvalues μ_k as input parameters. In order to solve it, we introduce the operator

$$T : \mathbb{R}_+^v \times \mathbb{R}_+^v \times \mathbb{R}_+ \to \mathbb{R}_+^v, \qquad T_k(\lambda;\mu;\rho) = \mu_k \frac{\alpha}{\alpha_k}(\rho;\lambda), \qquad k = 1,\ldots,v. \tag{5}$$

and recast eq. (4) in the equivalent vectorial form $\lambda = T(\lambda;\mu;\rho)$. Hence, we see that the full eigenvalue spectrum λ is a fixed point of the operator T. This suggests to obtain λ as the limit of a sequence

$$\lambda^{(0)} = \mu, \quad \lambda^{(n+1)} = T(\lambda^{(n)};\mu;\rho); \qquad\qquad \lambda = \lim_{n\to\infty} \lambda^{(n)}. \tag{6}$$

Under proper conditions on ρ and μ, the sequence can be shown to converge, see ref. [2]. The convergence is proved by showing that *i*) the sequence is component–wise monotonic increasing; *ii*) the sequence is component–wise bounded from above by any fixed point of T; *iii*) if T has a fixed point, this must be unique. Since λ is a fixed point of T, the first two properties imply that the sequence converges. Moreover, it can be easily shown that the limit is a fixed point of T. Property *iii*) guarantees that the limit of the sequence is precisely λ.

3 Numerical computation of the Gaussian integrals

The integrals α and α_k cannot be calculated with paper and pencil, owing to the symmetry mismatch between $\mathscr{N}(0,\Sigma)$ and $\mathscr{B}_\nu(\rho)$. However, it is possible to resort to numerical computation. Many years ago Ruben [1] has shown that α can be expanded as an infinite sum of chi–square distributions

$$\alpha(\rho;\lambda) = \sum_{m=0}^{\infty} c_m(s;\lambda) F_{\nu+2m}(\rho/s)\,, \tag{7}$$

with known coefficients c_m. The scale factor s has the same physical dimension as ρ and λ. It is introduced in order to factorize the dependence of α upon ρ and λ at each order of the expansion. We can prove that α_k admits an analogous series representation

$$\alpha_k(\rho;\lambda) = \sum_{m=0}^{\infty} c_{k;m}(s;\lambda) F_{\nu+2(m+1)}(\rho/s) \tag{8}$$

with different coefficients $c_{k;m}$. A detailed discussion of the coefficients c_m and $c_{k;m}$ is out of reach in this context. About this, we refer the reader to refs. [1, 2]. Eqs. (7) and (8) can be approximated with controlled systematic error by retaining a few terms. It is not difficult to achieve a fast and robust numerical implementation. In typical situations, the minimum number of terms required to keep the error below 1.0×10^{-14} does not exceed a hundred.

4 Numerical experiences

Given $\varepsilon > 0$, the number of steps $n_{\rm it}$ needed for an approximate convergence with relative precision ε, i.e.

$$n_{\rm it} \equiv \min_{n\geq 1}\left\{ n:\ \frac{||\lambda^{(n)}-\lambda^{(n-1)}||_\infty}{||\lambda^{(n-1)}||_\infty} < \varepsilon \right\}, \tag{9}$$

depends not only upon ε, but also on ρ and μ. In order to characterize distributionally the convergence rate of the reconstruction process, we must integrate out the fluctuations of $n_{\rm it}$ due to changes of μ, i.e. we must average $n_{\rm it}$ by letting

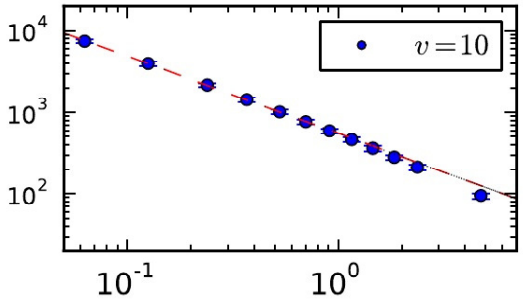

Fig. 1 Log–log plot of $\bar{n}_{\rm it}$ vs. ρ at $v = 10$ from a Monte Carlo simulation of $n_{\rm it}$ with λ taken from $\Sigma \sim \mathscr{W}_v(p, p^{-1}\mathbb{I}_v)$, $p = 2v$, $\omega = 1.9$ and $\varepsilon = 1.0 \times 10^{-7}$.

μ fluctuate across its own probability space. In this way, we obtain the quantity $\bar{n}_{\rm it} \equiv \mathbb{E}_\mu[n_{\rm it}|\varepsilon_{\rm T}, \rho]$, which better synthesizes the cost of the reconstruction for given ε and ρ. To proceed concretely, we extract λ from $\Sigma \sim \mathscr{W}_v(p, p^{-1}\mathbb{I}_v)$ (the ensemble of Wishart matrices with p degrees of freedom and scale matrix $p^{-1}\mathbb{I}_v$) with p properly chosen, and then we obtain μ by truncating λ according to eq. (4).

The basic iteration of eq. (6) is too slow to be of practical interest. For instance, at $v = 10$, $\rho \simeq 0.13$ and $\varepsilon = 1.0 \times 10^{-7}$ (corresponding to a reconstruction of λ with single floating–point precision) it is rather easy to extract realizations of μ which require $n_{\rm it} \simeq 15.000$ to converge. An improvement of the basic scheme is achieved via over–relaxation, i.e.

$$\begin{cases} \lambda_k^{(0)} = \mu_k \,, \\ \lambda_k^{(n+1)} = \lambda_k^{(n)} + \omega \left[T_k(\lambda^{(n)}) - \lambda_k^{(n)} \right] , \qquad k = 1, \ldots, v \end{cases} . \tag{10}$$

An example of dependence of $\bar{n}_{\rm it}$ upon ρ with the over–relaxed scheme is shown in Fig. 1. From the plot, we recognize the scaling law

$$\log \bar{n}_{\rm it}(\rho, v, \varepsilon) = a(v, \varepsilon) - b(v, \varepsilon) \log \rho \,, \tag{11}$$

showing that $\bar{n}_{\rm it}$ increases polynomially in $1/\rho$ and exponentially in v.

References

1. Ruben H. Probability content of regions under spherical normal distributions, IV: the distribution of homogeneous and non–homogeneous quadratic functions of normal variables. *The Annals of Mathematical Statistics*, 33(2):pp. 542–570, 1962.
2. Palombi F., Toti. S, Filippini R. Numerical reconstruction of the covariance matrix of a spherically truncated multinormal distribution. [arXiv:1202.1838 math.ST]